Human Molecular Genetics, 1998, Vol. 7, No. 12 **1967–1978**

Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the *t*-complex and under-representation on the X chromosome

Minoru S. H. Ko*, Tracy A. Threat, Xueqian Wang, Joseph H. Horton, Yushun Cui, Xiaohong Wang, Eric Pryor, Jason Paris, Jeannine Wells-Smith, John R. Kitchen, Lucy B. Rowe[1], Janan Eppig[1], Toshihiko Satoh[2], Larry Brant[3], Hiroyuki Fujiwara, Shinichi Yotsumoto and Hiroshi Nakashima

Center for Molecular Medicine and Genetics and Department of Internal Medicine, Wayne State University School of Medicine, 5047 Gullen Mall, Detroit, MI 48202, USA, [1]The Jackson Laboratory, Bar Harbor, ME 04609, USA, [2]Department of Hygiene and Public Health, Tokyo Women's Medical College, Tokyo 160, Japan and [3]Research Resources Branch, Gerontology Research Center, National Institute on Aging, NIH, Baltimore, MD 21224, USA

Received July 3, 1998; Revised and Accepted August 28, 1998

DDBJ/EMBL/GenBank accession nos 1051017 and 1054202

Mammalian embryos can only survive if they attach to the uterus (implantation) and establish proper maternal–fetal interactions. To understand this complex implantation pathway, we have initiated genomic analysis with a systematic study of the cohort of genes expressed in extraembryonic cells that are derived from the conceptus and play a major role in this process. A total of 2103 cDNAs from the extraembryonic portion of 7.5-day post-conception mouse embryos yielded 3186 expressed sequence tags, ~40% of which were novel to the sequence databases. Furthermore, when 155 of the cDNA clones with no homology to previously detected genes were genetically mapped, apparent clustering of these expressed genes was detected in subregions of chromosomes 2, 7, 9 and 17, with 6.5% of the observed genes localized in the *t*-complex region of chromosome 17, which represents only ~1.5% of the mouse genome. In contrast, X-linked genes were under-represented. Semi-quantitative RT–PCR analyses of the mapped genes demonstrated that one third of the genes were expressed solely in extraembryonic tissue and an additional one third of the genes were expressed predominantly in the extraembryonic tissues. The over-representation of extraembryonic-expressed genes in dosage-sensitive autosomal imprinted regions and under-representation on the dosage-compensated X chromosome may reflect a need for tight quantitative control of expression during development.

INTRODUCTION

A recent survey of mouse gene knockouts and spontaneous mutations concludes that implantation and subsequent development of the placenta are critical for the survival of the mammalian embryo *in utero* (1). In the human, clinicians have observed that about one to two thirds of human embryos die before birth, with the majority of these deaths occurring in the peri-implantation period (2,3). The key to the maternal–fetal interaction is the extraembryonic cells, which are derived from the conceptus but do not participate in formation of the embryo proper (1,4,5). Among these cells, trophoblast giant cells especially play multiple important roles as: (i) the leading edge of embryo invasion of the maternal endometrium; (ii) an immunological barrier protecting the embryo from the maternal immune response; and (iii) an endocrine organ, synthesizing and secreting steroid and peptide hormones (4–6).

Genomic imprinting plays an important role in the formation and function of extraembryonic tissues (7,8). Androgenotes carrying two paternally derived haploid genomes develop a relatively normal extraembryonic trophoblast and yolk sac, whereas the embryo itself develops very poorly (9,10). Complete human hydatidiform moles, for example, are androgenotes (11). In contrast, gynogenotes carrying two maternally derived haploid genomes show relatively normal development of the embryo itself, but the extraembryonic tissues develop poorly (9,10). In another example of parental-specific dosage control, the

*To whom correspondence should be addressed. Tel: +1 313 577 6708; Fax: +1 313 577 6200; Email: msko@cmb.biosci.wayne.edu

paternally derived X chromosome is preferentially inactivated in extraembryonic tissues of placental mammals (12), whereas in other somatic tissues X chromosome inactivation occurs randomly (13).

As a first step toward understanding the function of extraembryonic tissues at the implantation site, we employed systematic sequencing and mapping analyses of genes expressed in the upper half (extraembryonic portion) of the 7.5 days post-conception (d.p.c.) mouse embryo. The study reveals many interesting genes, many of them unexpectedly clustered in the mouse genome. Possible implications of this finding for extraembryonic tissue function and genomic imprinting are discussed.

RESULTS

Construction of a cDNA library and profiles of collected cDNAs

We constructed a cDNA library from the upper half of the 7.5 d.p.c. mouse embryo. Although the tissue consists of trophoblast giant cells, the ectoplacental cone (EPC, precursor of trophoblast cells, the largest mass in the sample) and the extraembryonic ectoderm (Fig. 1), for brevity this cDNA library will be called the 7.5 EPC library and the genes recovered from this library will be called 7.5 EPC genes. A total of 2103 cDNAs picked randomly from this cDNA library were sequenced to develop 3186 expressed sequence tags (ESTs) (1989 3′-ESTs, 1197 5′-ESTs). Comparisons among these ESTs suggest that there are 751 distinct cDNAs in this group. Sequence searches of these 3′-ESTs against a public database using Blastn (14) revealed that 40% (804/1989) of the cDNA species were previously unknown in the non-redundant GenBank database (which contains annotated DNA sequence information). Furthermore, searches against the public EST database (dbEST) (15) also showed that a significant fraction (17%, 329/1989) of ESTs from the 7.5 EPC library were new. Thus it appears that many developmentally important genes have not been detected in the current large-scale EST projects, very likely because of tissue bias (16).

The frequency of individual cDNA clones included in this EST collection is a reflection of the expression level of these genes in the tissue source used for cDNA library construction (17). Table 1 shows the gene expression profiles of 3′-ESTs which were identified as known genes listed according to the frequency with which they were found in this cDNA library. Well-recognized cDNA sequences with characteristic functions in extraembryonic tissue at this developmental stage were recovered at levels appropriate for the extraembryonic tissues: for example, prolactin-like protein A (*Prlpa*, 0.65% of the total), placental lactogen-I (*Pl1*, 0.60%), H19 (*H19*, 0.50%), proliferin (*Plf*, 0.40%), placental lactogen-I variant (RNU32679, 0.15%), Cea5 (MUSCEA5, 0.15%), connexin31 (MMCX31, 0.10%), endo B cytokeratin (MUSKTCEB, 0.10%), placenta and embryonic expression gene (*Pem*, 0.10%) and X-inactivation-specific transcript (*Xist*, 0.05%). Genes were also encountered which had not previously been shown to be expressed in extraembryonic tissues, e.g. adrenomedullin (*Adm*, 0.35%), *Wnt6* (0.10%) and tropomyosin-4 (RNTM4, 0.10%). The presence of these genes at relatively high abundance suggests that they have a previously unrecognized importance in extraembryonic tissue development. Indeed, we have recently demonstrated by *in situ* hybridization and northern blot analyses that the *Adm* gene, which encodes a potent hypotensive peptide, is highly expressed in trophoblast giant cells at the implantation site (18).

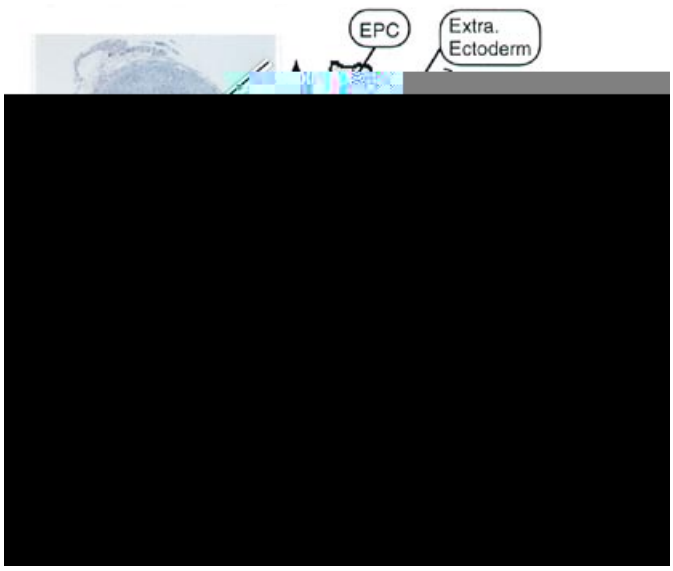

Figure 1. Experimental strategy. (Left) Transverse section through entire uterus showing implanted 7.5 d.p.c. mouse embryo. The section is stained with hematoxylin and eosin. (Right) Schematic representation of the embryo. EPC, ectoplacental cone; Extra. Ectoderm, extraembryonic ectoderm; TGC, trophoblast giant cells.

Interestingly, the expression profile also includes a group of genes previously detected as expressed in vascular endothelial cells, e.g. *Adm* (0.35%) and VE-cadherin (MMVECADH, 0.10%). This suggests some similar properties of extraembryonic trophoblast giant cells and vascular endothelial cells. It is consistent with a recent report that suggests such a similarity based on a comparable switch in integrin expression (19), even though the embryological origins of the two cell types are completely different (5,20).

Genetic mapping of newly identified cDNAs

Localization of new genes on the mouse genetic map was performed as described (21). In brief, unique PCR primer pairs were designed from the 3′-untranslated region (UTR) of each novel cDNA. The use of the 3′-UTR of cDNAs as the target for PCR amplification increases the probability of finding sequence polymorphisms between C57BL/6J and *Mus spretus* (22). A total of 751 PCR primer pairs were synthesized. Approximately 500 of the primer pairs gave a unique PCR product of the expected size from both mouse strain DNAs in our standardized PCR assay, a success rate in line with previous results based on 3′-UTR sequences from other cDNA sequences from GenBank as the source of markers (21).

The original gene mapping method used primarily PCR–RFLP (22), but we have recently noted that ~35% of usable primer pairs can detect polymorphism between C57BL/6J and SPRET/Ei (*M.spretus*) by simple heteroduplex analysis (23). Most of the primer pairs described in this paper were mapped by this heteroduplex method. These primer pairs were used to type an interspecific backcross mouse panel (BSS panel of 94 animals [(C57BL/6JEi × SPRET/Ei) × SPRET/Ei]) from the Jackson Laboratory (24). The data were analyzed using the Map Manager program (25).

Table 1. Genes that appeared more than twice in the 3′-EST collection

Identification	Frequency	Identification	Frequency
Mus musculus breast heat shock 73 protein (hsc73) mRNA	16	*Homo sapiens* mRNA for gCIq-R	2
Mus musculus protein synthesis elongation factor Tu (eEF-Tu)	14	*Haemophilus influenzae* era, fdhE, holD, lepA, lepB, rimI	2
Rat prolactin-like protein A (rPLP-A) mRNA	13	*Homo sapiens* splicing factor, arginine/serine-rich 7 (SFRS7)	2
Mouse placental lactogen I (PL-I) mRNA	12	*Mus domesticus* fetus cerebral cortex mRNA	2
Mus musculus H19 mRNA	10	*Mus domesticus* mRNA for 14-3-3 ε	2
Mouse ferritin light chain	10	Human capping protein α mRNA	2
Mouse heat shock protein 86 mRNA	9	Human cytochrome bc-1 complex core protein II mRNA	2
Mouse cytoplasmic γ-actin gene	8	Human cytochrome c1 mRNA	2
Mouse proliferin mRNA	8	Human DNA sequence from cosmid V870H8	2
Mouse (clone M1) GTPase (Ran) mRNA	7	Human fetal brain: promyelogeneous leukocyte mRNA for TPRD	2
Mouse cytoskeletal mRNA for β-actin	7	Human HLA-B-associated transcript 3 (BAT3) mRNA	2
Mouse mRNA for translational controlled 40 kDa polypeptide	7	Human hnRNP A2 protein mRNA	2
Human mRNA for ORF	6	Human mRNA (KIAA0071) for ORF (novel protein)	2
Mus musculus mRNA for testis-specific thymosin β-10	6	Human mRNA for 19 kDa protein of signal recognition particle	2
Mouse 84 kDa heat shock protein mRNA	6	Human mRNA for KIAA0183 protein	2
Mouse glyceraldehyde-3-phosphate dehydrogenase mRNA	6	Human mRNA for ORF	2
Mouse nucleolin gene	6	*Mus musculus* gene for α-catenin	2
Rat major α-globin mRNA	6	*Mus musculus* AP-2.2 gene	2
Rat mitochondrial ATP synthase β subunit mRNA	6	*Mus musculus* bgbp gene for β-galactoside binding protein	2
Rat mRNA for adrenomedullin precursor	6	*Mus musculus* DNA for α globin gene and flanking regions	2
Mus musculus 45S pre-rRNA gene	5	*Mus musculus* gene for haem oxygenase (exon 5)	2
Mus musculus mRNA for ribosomal protein S3	5	*Mus musculus* mitochondrial 12S rRNA	2
Mus musculus mRNA for tropomyosin 5 (3′-UTR)	5	*Mus musculus* mitochondrial malate dehydrogenase gene	2
Mouse 18S rRNA gene 5′-end	5	*Mus musculus* mRNA for connexin31	2
Mouse α-tubulin gene M-α-2, 3′-end	5	*Mus musculus* mRNA for L14 lectin	2
Mouse metallothionein II (MT-II) gene	5	*Mus musculus* mRNA for Rab7 protein	2
Mouse mRNA for elongation factor 1-α (EF 1-α)	5	*Mus musculus* mRNA for ribosomal protein L37a	2
Mouse mRNA for ribosomal protein S6	5	*Mus musculus* mRNA for ribosomal protein L5, 3′-end	2
Mouse nucleolar protein N038 mRNA	5	*Mus musculus* mRNA for secretin	2
Mus musculus calmodulin synthesis (CaM) cDNA	5	*Mus musculus* mRNA TNZ2 for p68 RNA helicase	2
Mus domesticus adult brain mRNA	4	*Mus musculus* NRF1 mRNA	2
Human initiation factor eIF-5A gene	4	*Mus musculus* rpS6 gene	2
Human mRNA (KIAA0045) for ORF	4	*Mus musculus* VE-cadherin gene (λ5 clone)	2
Mus musculus mRNA for poly(A)-binding protein	4	Mouse argininosuccinate synthetase (Ass) mRNA	2
Mus musculus mRNA for TAX-responsive element-binding protein	4	Mouse calmodulin (Cam I) mRNA	2
Mouse A-X actin mRNA	4	Mouse cyclophilin mRNA	2
Mouse catalase mRNA	4	Mouse elongation factor 2 (ef-2) mRNA, 3′-end	2
Mouse heat shock protein hsp84 mRNA	4	Mouse embryonal carcinoma cell mRNA for 3-hydroxyacyl CoA	2
Mouse mRNA for acidic ribosomal phosophoprotein PO	4	Mouse endo B cytokeratin mRNA	2
Mouse mRNA for G protein β subunit homolog	4	Mouse gelsolin gene	2
Mouse mRNA for nuclear pore-targeting complex component	4	Mouse insulinoma (rig) mRNA	2
Mouse ribosomal protein S4 (Rps4) mRNA	4	Mouse integrin β4 subunit mRNA	2
Murine mRNA for J1 protein, yeast ribosomal protein L3	4	Mouse LT lymphotoxin (LT) gene	2
Mus musculus acidic ribosomal phosphoprotein P1 mRNA	4	Mouse mRNA for α-enolase (2-phospho-d-glycerate)	2
Mus musculus CDC42 mRNA	4	Mouse mRNA for apg-2	2
Mus musculus ERCC2 gene	4	Mouse mRNA for CArG-binding factor-A	2
Mus musculus fat facets homolog (Fam) mRNA	4	Mouse mRNA for cysteine-rich glycoprotein SPARC	2
Rattus norvegicus mRNA for ribosomal protein L4	4	Mouse mRNA for monoclonal non-specific suppressor factor	2
Rattus norvegicus mRNA for ribosomal protein L41	4	Mouse mRNA for pyrubate kinase M	2
Rattus rattus mRNA for ribosomal protein L11	4	Mouse placenta and embryonic expression gene (pem) mRNA	2
Rattus rattus mRNA for ribosomal protein L23a	4	Mouse ribosomal protein L19	2
Rat mRNA for ribosomal protein L17	4	Mouse surfeit locus (Surf-3) processed pseudogene 2	2

Continued overleaf

Table 1. *Continued*

Identification	Frequency	Identification	Frequency
Rat mRNA for ribosomal protein L18a	4	Mouse ubf gene for transcription factor UBF	2
Rat mRNA for ribosomal protein L7a	4	Mouse Wnt-6 mRNA	2
Sequence 5 from US patent 4758511	3	*Mus musculus* (clone pVZmSin3B) mSin3B mRNA	2
Cricetulus longicaudatus rrs1 gene for arginyl-tRNA synthetase	3	*Mus musculus* (clone R24) rds gene	2
Homo sapiens mRNA for activin β-C chain	3	*Mus musculus* (clone: pMAT1) mRNA	2
Homo sapiens HnRNP F protein mRNA	3	*Mus musculus* α-globin mRNA	2
Human histone H2A.Z gene, upstream promoter sequence	3	*Mus musculus* B6D2F1 clone 2C#4 mRNA	2
Leucine aminopeptidase (cattle, kidney, mRNA, 2056 nt)	3	*Mus musculus* chaperonin mRNA	2
Mus musculus mRNA for calmodulin	3	*Mus musculus* common cytokine receptor γ chain gene	2
Mus musculus mRNA for mitochondrial gene for subunit I	3	*Mus musculus* cytokeratin no. 19 mRNA	2
Mus musculus syndecan-1	3	*Mus musculus* folate-binding protein gene, 5′-end	2
Mus musculus tdgf1 gene	3	*Mus musculus* Hsp70-related NST-1 (hsr.1) mRNA	2
Mouse 18S, 5.8S, 28S rRNA gene cluster (clone pMEB3)	3	*Mus musculus* LAF1 transketolase mRNA	2
Mouse α-tubulin isotype M-α-6 mRNA	3	*Mus musculus* laminin α5 chain (Lama5) mRNA	2
Mouse ATP synthase α subunit	3	*Mus musculus* lymphoid-specific transcription factor NFATc3	2
Mouse ERp99 mRNA encoding an endoplasmic reticulum	3	*Mus musculus* MT transposon-like element, clone MTi7	2
Mouse facilitated glucose transport protein mRNA	3	*Mus musculus* nuclear-encoded mitochondrial acyltransferase	2
Mouse Ig-related glycoprotein-70 mRNA	3	*Mus musculus* Paneth cell enhanced expression PCEE mRNA	2
Mouse lactate dehydrogenase A-4 (LDH-A) mRNA	3	*Mus musculus* phosphoglycerate kinase (Pgk1-ps1) processed	2
Mouse mRNA for cyclophilin (EC 5.2.1.8)	3	*Mus musculus* retinoblastoma-binding protein (mRbAp46) mRNA	2
Mouse mRNA for prothymosin α	3	*Mus musculus* ribosomal protein (Ke-3) mRNA	2
Mouse mRNA for ubiquitin	3	*Mus musculus* ribosomal protein L9 (musl9) mRNA	2
Mouse replication-dependent histone H2A.1 gene	3	*Mus musculus* translation initiation factor 4E (eIF-4E)	2
Mus musculus Cea5 mRNA	3	*Mus* sp. nucleic acid-binding protein mRNA	2
Mus musculus large ribosomal subunit protein mRNA	3	Phosphoglycerate mutase type B subunit (rats, liver, mRNA)	2
Mus musculus ribosomal protein L8 (RPL8) mRNA	3	*Rattus norvegicus* (Sprague–Dawley) ribosomal protein S23 mRNA	2
Mus musculus ribosomal protein S26 (RPS26) mRNA	3	*Rattus norvegicus* C15 mRNA	2
Mus musculus RNA helicase mRNA	3	*Rattus norvegicus* genes for 18S, 5.8S and 28S rRNAs	2
Mus musculus translation initiation factor (Eif4g2) mRNA	3	*Rattus norvegicus* mRNA for elongation factor 1α	2
Rattus norvegicus (Sprague–Dawley) ARL5 mRNA for ARF-like protein	3	*Rattus norvegicus* mRNA for ribosomal protein S9	2
Rattus norvegicus mRNA for ribosomal protein S21	3	Rat alternative brain Ca^{2+}-ATPase mRNA	2
Rat brain Ca^{2+}-ATPase mRNA	3	Rat clathrin heavy chain mRNA	2
Rat mRNA for β-globin	3	Rat liver mRNA for proteasomal ATPase (TBP1)	2
Rat mRNA for ribosomal protein S3	3	Rat mitochondrial proton/phosphate symporter mRNA	2
Rattus norvegicus clone C426 intestinal epithelium	3	Rat mRNA for 14-3-3 protein γ-subtype	2
Rattus norvegicus placental lactogen-I variant mRNA	3	Rat mRNA for ribosomal protein L23	2
42 kDa membrane glycoprotein (human, gastric carcinoma cells)	2	Rat mRNA for ribosomal protein L35	2
β-HKA, H,K-ATPase β-subunit (rat)	2	Rat mRNA for ribosomal protein S10	2
Bovine mitochondrial adenylate kinase (AK2B) mRNA	2	Rat mRNA for ribosomal protein S14	2
Bovine mRNA for actin2	2	Rat mRNA for tissue factor pathway inhibitor	2
Chinese hamster gene for polyubiquitin	2	Rat ribosomal protein L18 mRNA	2
Chinese hamster ovary (CHO) ribosomal protein S14 mRNA	2	Rat TM-4 gene for fibroblast tropomyosin 4	2
Chinese hamster P1 protein mRNA	2	*Rattus norvegicus* eukaryotic initiation factor 5 (eIF-5)	2
Cricetulus griseus β-tubulin isotype I mRNA	2	*Rattus norvegicus* gene 33 DNA, exon 4 and 3′-end	2
Cyclotella spp. rRNA large subunit	2	*Rattus norvegicus* taipoxin-associated calcium-binding protein	2
Differentially expressed homolog of HepG2 cDNA clone	2	*Rattus rattus* eukaryotic initiation factor (Eif-2) 67 kDa	2
Homo sapiens mitoxantrone-resistance associated mRNA	2	SCP-2, sterol carrier protein-2 {3′-region} (mice, C57BL/6)	2

The locations of 155 new cDNAs are shown in Figure 2. For clarity of presentation, the map shows only the locations of genes derived from this 7.5 EPC library, although the map is anchored extensively by 354 Mit markers and many gene markers through the contributions of many researchers (24).

Unexpectedly, these novel 7.5 EPC genes showed an uneven distribution in the mouse genome (Fig. 2). Genes were apparently clustered in subregions of many chromosomes: chromosome 2 (proximal and distal), chromosome 4 (distal), chromosome 5 (proximal), chromosome 7 (proximal and distal), chromosome 9

Figure 2. Genetic map of the Jackson Laboratory BSS cross (data taken from July 1997 release) showing the distribution of 7.5 EPC genes. The map shows all positions that contain loci marked with ticks and lists the specific locations of genes derived from the 7.5 EPC library. The map was anchored by 354 Mit markers and many gene markers through the contributions of many researchers (see http://www.jax.org/resources/documents/cmdata/bkmap/BSS.html).

(middle), chromosome 11 (distal) and chromosome 17 (proximal) (Fig. 2).

Statistical analysis of the distribution of 7.5 EPC genes on the mouse genome

Because of the involvement of the X chromosome in extraembryonic tissue function (see Introduction), we originally hypothesized that many genes would map to the X chromosome. The results were quite the opposite. The probability of appearance of the 7.5 EPC genes on the X chromosome was compared with that of a reference population of all coding loci (2504) on the backcross panel. The reason for using all coding loci as a reference population (rather than the physical length of the map) is that distribution of genes in general is not uniform in the mouse genome (26). This sample should be relevant for both autosomes and the X chromosome and would correct for any chromosome-specific bias, such as a lower density of genes, as reported for the human X chromosome (27). As shown in Table 2, 4.7% (118/2504) of the reference population were localized on the X chromosome, while 1.3% (2/155) of the 7.5 EPC genes were localized on the X chromosome. To test the hypothesis, the odds ratio (OR) and the 95% confidence interval (95% CI) were calculated. Contrary to our original expectation, the X chromosome was significantly under-represented in 7.5 EPC genes (OR 0.264, 95% CI 0.031, 0.994).

Secondly, to investigate those chromosomes (or regions) where 7.5 EPC genes are abundantly distributed compared with the reference population, χ^2 tests of 2×2 tables were conducted for each chromosome (Table 2). The Bonferroni correction was employed for this exploratory analysis, to avoid multiple testing problems. With the Bonferroni correction, a *P*-value of <0.0023 (0.05/22) was considered as statistically significant. By this criterion, chromosome 17 showed a higher density of EPC-expressed genes (P = 0.003). The proximal domain of chromosome 17 includes the *t*-complex region (28,29) and, when further statistical analysis was conducted by splitting chromosome 17 into *t*-complex (centromere to *H2-M2*) and non-*t*-complex regions, statistically significant clustering of 7.5 EPC genes in the *t*-complex was observed (P = 0.0005). In addition to the novel genes, we found three other previously defined genes (*Tcp1*, *H2-Ke2* and *H2-Ke4*) in the cDNA collection that also map to this region (Fig. 3B).

These results clearly indicate clustering of the 7.5 EPC genes on the mouse *t*-complex.

Table 2. Comparison of chromosomal distribution of markers between all genetic loci and 7.5 EPC gene loci on the BSS backcross map

Chromosome	All loci		Wsu loci		*P*-value
	No. of loci	%	No. of loci	%	
1	156	6.2	5	3.2	0.128
2	201	8.0	14	9.0	0.656
3	112	4.5	7	4.5	0.988
4	163	6.5	11	7.1	0.774
5	162	6.5	11	7.1	0.759
6	127	5.1	9	5.8	0.687
7	161	6.4	11	7.1	0.743
8	137	5.5	7	4.5	0.610
9	99	4.0	10	6.5	0.128
10	115	4.6	8	5.2	0.744
11	179	7.1	9	5.8	0.527
12	82	3.3	3	1.9	0.483
13	146	5.8	7	4.5	0.496
14	94	3.8	3	1.9	0.373
15	80	3.2	7	4.5	0.370
16	86	3.4	6	3.9	0.773
17	101	4.0	14	9.0	0.003
17 (*t*-complex)	52	2.1	10	6.5	0.0005
17 (outside *t*)	49	2.0	4	2.6	0.590
18	84	3.4	5	3.2	0.931
19	94	3.8	6	3.9	0.941
X	118	4.7	2	1.3	0.045
Y	7	0.3	0	0.0	>0.999
Total	2504	100.0	155	100.0	

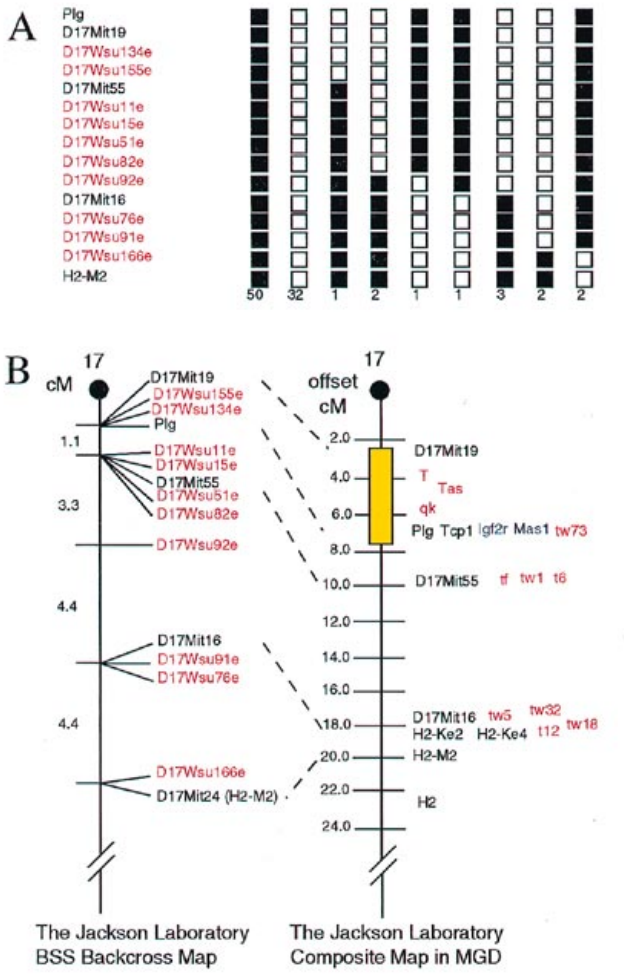

Figure 3. (**A**) Haplotype figure of the Jackson Laboratory BSS proximal chromosome 17 mapping data. Loci are listed in order with the most proximal at the top, with no order implied for co-segregating loci. The filled boxes represent the C57BL/6J allele and the open boxes represent the SPRET/Ei allele. Numbers at the bottom of each column of boxes are the total number of animals with that haplotype. (**B**) Detailed map of the Jackson Laboratory BSS chromosome 17 proximal region (left) compared with the composite map with important landmarks (right). (Left) Red, Wsu markers; black, anchor markers. (Right) red, mouse mutant loci; black, anchor markers; blue, known imprinted genes.

Expression analyses of mapped genes by RT–PCR

To test whether these mapped genes are expressed specifically in the extraembryonic tissues, we performed semi-quantitative RT–PCR analyses of genes mapped in this study. The same amount of cDNA pools derived from extraembryonic tissues and embryonic tissues were sequentially diluted in 10-fold steps. A cDNA fragment was amplified from each diluted cDNA pool by PCR with gene-specific primer pairs and run on non-denaturing polyacrylamide gels. Figure 4 shows an example of the ethidium bromide stained gels. The amount of each gene-specific cDNA fragment contained in the individual cDNA pools can be roughly estimated by their relative band intensities in the gel. As a control, the abundance of cytoskeletal β-actin cDNA in the extraembryonic tissues was shown to be the same as that in the embryonic tissues. In contrast, the adrenomedullin (*Adm*) gene, which has been shown to be expressed specifically in extraembryonic tissues (especially in trophoblast giant cells; 18), showed PCR products only in extraembryonic tissues. Thus, limiting dilution RT–PCR analysis appears to reflect the expression levels of genes both in extraembryonic and embryonic tissues.

Data obtained for 150 cDNA species are summarized in Figure 5. Of the genes examined by RT–PCR, 32.7% (49/150) were expressed only in the extraembryonic tissues, 9.3% (14/150) were expressed in the extraembryonic tissue at levels 100-fold higher than in embryonic tissue and 24.0% (36/150) were expressed in the extraembryonic tissue at levels 10-fold higher than in embryonic tissue.

Because we have not selected cDNA clones which are uniquely expressed in the extraembryonic tissues, this cDNA collection should and does also include genes expressed in embryonic tissues. For example, 17.3% (26/150) of genes showed the same level of expression both in extraembryonic and embryonic tissues and 2.7% (4/150) showed 10-fold higher expression in the embryonic tissues than in the extraembryonic tissues. Furthermore, 12.7% (19/150) of genes were expressed at a very low level in both embryonic and extraembryonic tissues and yielded no RT–PCR products and 1.3% (2/150) of genes showed only embryonic expression. The lack of detectable expression of these genes in extraembryonic tissues may mean that some cDNA clones with very low expression levels were included by chance. Alternatively, these clones may be the result of slight contamination of the tissue sample used to make the clones with some extraembryonic cells.

In summary, about two thirds of the genes showed either unique expression in the extraembryonic tissues or higher expression in extraembryonic than in embryonic tissues. This confirms the expectation that ESTs sampled from the extraembryonic tissues reflect the gene expression patterns in those tissues.

Characterization of genes clustered in the *t*-complex region

Where are the new 'clustered' genes within the *t*-complex and are they actually related in any way? Figure 3 compares a magnified view of the genetic map of the chromosome 17 proximal region with the composite map, anchored by important landmarks. Genes expressed in the extraembryonic portion of the 7.5 d.p.c. embryo were clustered in five subregions of the *t*-complex (Wsu loci in red). The most proximal bin may result from the effects of the large inversion, 'Inversion 2', between C57BL/6J and *M.spretus* (29). As is well documented, inversion reduces observed meiotic recombination in the region and causes distortion of the genetic map (29). However, this is the only known inversion between C57BL/6J and *M.spretus* and, thus, the overall clustering of genes on the genetic map may instead reflect their physical proximity. In support of this possibility, when a BAC library was screened using four of the cDNA clones chosen randomly as probes, we found that at least two genes, *D17Wsu11e* and *D17Wsu15e*, were recovered in a single 175 kb BAC clone. Southern blot analysis refined the distance between these two genes to <10 kb (data not shown).

Because the mapped 7.5 EPC genes showed no significant sequence similarity among themselves, clustering obviously does not reflect descent from a duplicated ancestral gene (30), as with

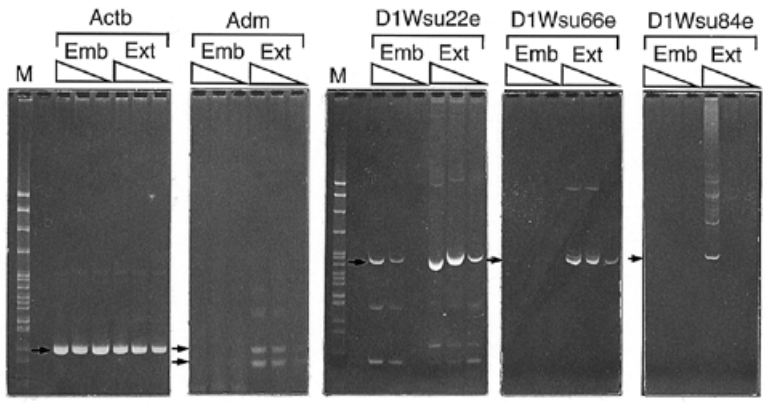

Figure 4. Example of RT–PCR analyses.

the *Hox* (31) and *globin* (32) gene clusters. To look more carefully for any sequence similarities, the 7.5 EPC cDNA clones that mapped to the *t*-complex were completely sequenced and no sequence motifs in common were found. Consistent with these results again, northern blot analyses also showed distinctive mRNA sizes and tissue distributions for each of them (data not shown). Taken together, these results suggest an unexpected *t*-complex clustering of many distinctive genes expressed in extraembryonic tissues.

DISCUSSION

At 7.5 d.p.c., dramatic developmental changes in embryonic tissues are laying down both organ anlage and the complex placental nourishment of the nascent fetus. At first sight it may nevertheless be unexpected that a considerable fraction of the genes recovered in this study were previously undetected and that they tend to cluster at some locations. Here we summarize the major findings and discuss the possible relationship of localization and regulation of cohorts of developmental genes.

Three outcomes of the study

Sequencing and mapping tissue-specific gene cohorts has become a standard and powerful entrée to systematic functional analysis. Here, with 3186 ESTs developed from the 7.5 d.p.c. extraembryonic zone and 155 previously undetected species mapped to 1 cM resolution in the genome, there were three significant outcomes.

(i) A significant fraction of the EST set showed no hits when compared with the 1 400 000 entries in dbEST. This suggests that genes specifically expressed during development are not found in the current large-scale EST projects, which have concentrated on mature tissues. Especially because of the technical and ethical/legal barriers to the study of human embryos, the mouse provides a key source of discovery of many genes uniquely important in the early development of placental mammals. Consequently, specialized EST projects like this one can aid in studies that range from positional cloning of disease genes to 'cyberscreening' for genes of interest (33). An indication of the potential usefulness of the 7.5 d.p.c. cDNA collection comes from the recent successful hunt for the Peutz–Jeghers syndrome gene (34), which found the only EST match in the public databases, a protein kinase, among the set described here.

(ii) EST sequence data, now deposited in the NCBI dbEST (15) adds expression profile information for randomly selected cDNA clones from the early embryonic tissue. Information about the relative expression of genes in tissues is thus obtained without northern analyses.

(iii) Systematic mapping of the genes provides a genome-wide view of the distribution of genes expressed in the tissue, with some hints of evolutionary and co-regulatory processes that are discussed further below.

Uneven genomic distribution of 7.5 EPC expressed genes

By RT–PCR analysis, one third of the cDNAs were expressed in both embryonic and extraembryonic compartments and likely include housekeeping genes, but another one third were expressed only in extraembryonic tissue and the rest were expressed predominantly in the extraembryonic zone. These results substantiate both the efficacy of the dissection used to define the EPC and the developmental specificity of gene expression.

Figure 4 color codes the level of expression of a number of the genes and places them at their mapped chromosomal locations for comparison. The striking gene clusters are apparent.

Clustering of coordinately expressed genes is of course common in bacterial operons (35). In mammalian genomes, however, gene clusters, where seen, are on a much larger scale (Mb rather than 5–20 kb) and may involve long-range chromatin interactions (36,37). Until now, two types of clusters have been seen. One, exemplified by the *Hox* (31) and *globin* (32) genes, is based on tandem duplication and subsequent divergent evolution of ancient lineages (30). A second, including the major histocompatibility complex (38) and sets of imprinted genes (7,39), involves genes that are coordinately regulated but structurally unrelated. The extent to which such clusters exist in mammals has been unknown, but this is, to our knowledge, the first report of a large cluster of genes whose defining common feature is their co-expression at a fixed time and location. Recently, however, Danieli *et al.* (40) have reported that genes expressed in skeletal muscle may be enriched on human chromosomes 17, 19, 21 and X. Thus, clustering could be part of the global structure of evolving chromosomes. On the other hand, the genes in the cohort examined here may be a collective exception to ordinary dispersal of most genes, based on the particular requirements of the developing embryo, discussed below.

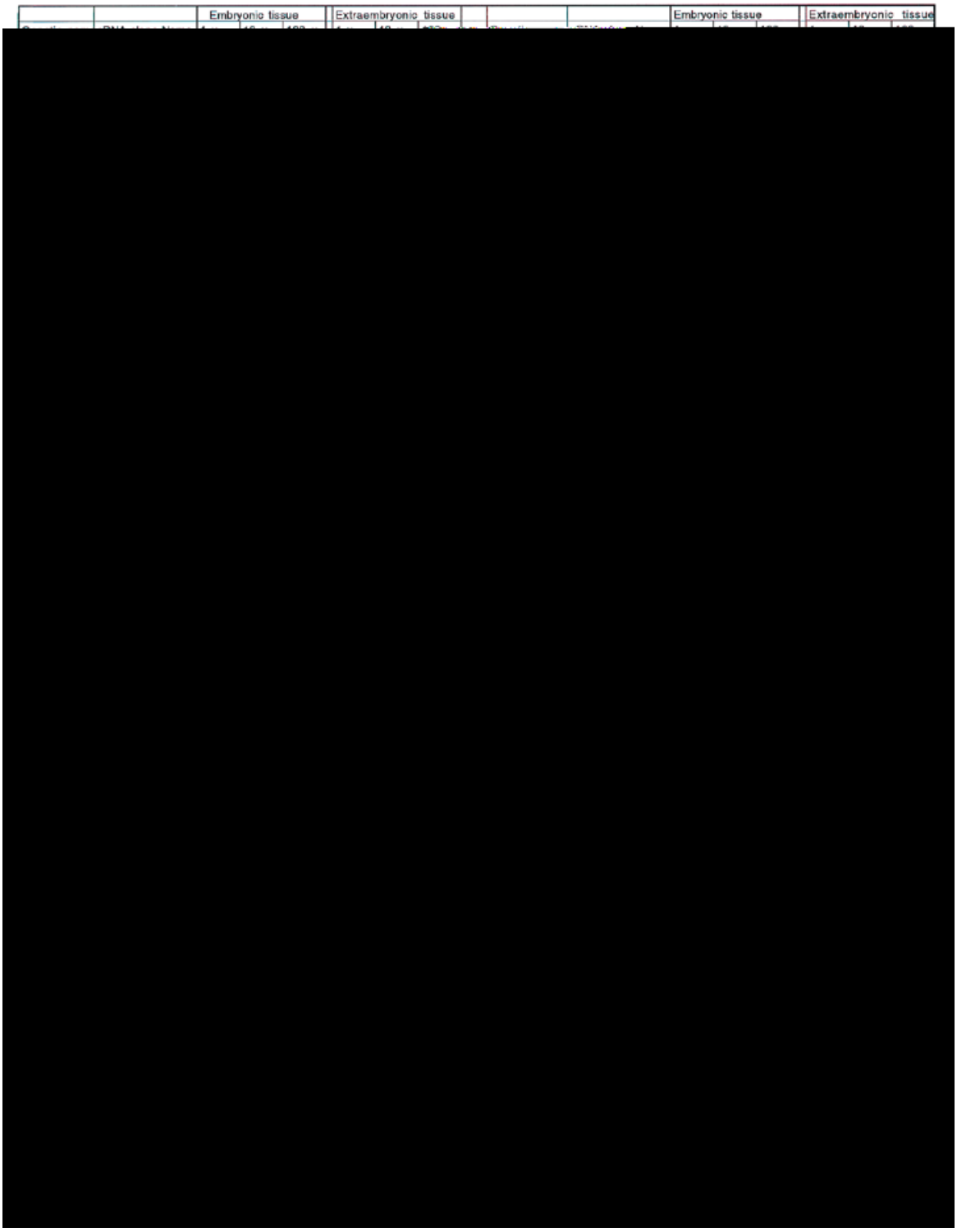

Figure 5. Summary of RT–PCR analyses.

Do dosage control and implantation requirements drive co-localization?

Possible mechanisms underlying clustering are suggested by features of the regions highlighted in this study and the key developmental functions of the genes involved.

First, the genes in regions like the *t*-complex are involved in functions like implantation, which are specific to placental mammals. In fact, all homozygous lethal mutants in the *t*-complex show abnormal phenotypes at or around the stage of implantation (28). An intriguing example is the t^{w73} mutant (t^{clw73}), which is known to affect implantation (41). Some of the

correlatively mapped new genes in Figure 4 become candidates for such *t*-lethal mutations.

Second, the regions that are enriched or deficient in EPC-expressed genes undergo tight regulation of gene dosage, by imprinting in the *t*-complex and other autosomal regions (7) and by dosage compensation on the X chromosome. Consistent with this possibility, apparent gene clusters co-localize with imprinted areas that include, in addition to the *t*-complex, proximal chromosome 2 and proximal and distal segments of chromosome 7 (Fig. 2; cf. ref. 39). The hint that these genes may be candidate imprinted loci is reinforced by the finding that *D17Wsu11e*, the first new gene to be tested in work in progress, shows hemizygous methylation (data not shown), a hallmark of genomic imprinting.

The *t*-complex also often exhibits haploinsufficiency in mutagenesis experiments (42). If a developmental pathway is particularly sensitive to dosage of the genes involved, this requirement might serve as a possible mechanism for haploinsufficiency (43). For example, nuclear transplantation experiments (9,10), analyses of complete hydatidiform moles (11) and analyses of embryos with uniparental disomy (44) demonstrate that, depending on the genomic segment involved, abnormal dosage of genes in imprinted regions results in overgrowth or deficient growth of extraembryonic tissues. One known imprinted gene, *Mash2*, specifically controls the growth of the placenta (45). In a formally similar manner, the extraembryonic tissues of mouse embryos with supernumerary X chromosomes develop poorly, resulting in death of the embryo *in utero* (46,47), and at least one locus on the X chromosome regulates the size of the placenta in interspecific crosses (48).

Thus, stringent control of extraembryonic tissue growth may be so critical for placental mammals that many genes expressed in extraembryonic tissue are localized in evolutionarily advantageous, dosage-sensitive genomic regions. The need for tight dosage control might also apply to growth of the embryo proper, because embryonal overgrowth is seen in mouse and human when the dosage of certain genes in the imprinted regions is increased or when a particular growth suppressing locus on the X chromosome is mutated (for human cases see refs 49,50).

Dosage control of genes included in these genomic regions provides additional options by control of the developmental time at which dosage is adjusted and on parental-specific contributions to the active gene complement. We note that there is an interesting coincidence of timing of the reimposition of imprinting and X-inactivation at this very stage of development of the EPC, which may have implications for the relationship between selective inactivation of genes and the implantation process. For imprinted genes, the experiments on nuclear transplantation (see above) have also shown that paternal alleles and maternal alleles are differentially obligatory for development of the embryo (9–11). Correlatively, X-inactivation is random in the somatic cells of eutherian mammalian embryos (13), but is uniquely paternal in the extraembryonic tissue (12). This pattern of parental contributions has given rise to considerable speculation about conflicts between maternal and paternal genomes during embryonic development (51). It now seems that the divergent roles of the two parental genomes may have evolved as an effective method for the specific control of gene expression in the regions where genes expressed in the cDNA collections reported here are over- or under-represented and that some of these genes may be mediators of this type of regulation.

MATERIALS AND METHODS

cDNA library and screening

Ten embryos at 7.5 d.p.c. were collected from timed pregnant C57BL/6J inbred mice (Jackson Laboratory) according to a standard procedure (52). Embryos were cut in two to separate the upper portion (ectoplacental cone and extraembryonic ectoderm) from the lower portion (embryonic ectoderm). Total cell RNA was extracted using a kit (Stratagene). Construction of the cDNA library was performed using a kit (Life Technologies) with ~50 ng total RNA mixed with 100 ng oligo(dT)-*Not*I primer (5′-AATTCGCGGCCCGCTTTTTTTTTTTTTTT-3′) for first strand cDNA synthesis in a 20 µl reaction. The cloning vector was pSPORT1 (Life Technologies) and cloning sites were *Sal*I and *Not*I. Cloning of inserts of average size 1.4 kb was carried out without normalization (53), in order to provide information about the abundance of the individual cDNA species.

Sequencing and sequence analyses

Plasmid DNA from arrayed cDNAs was isolated using the QIAwell-8 System (Qiagen) on 3 ml LB overnight cultures. Eluted DNA (150–200 ng/µl) was used directly in cycle sequencing reactions. Sequencing reactions were performed using Applied Biosystems (ABI) M13 primers and PCR followed by analysis on an ABI 373A Automated DNA Sequencer. Approximately 400 bp of the 3′ and 5′ ends of cDNA clones were obtained.

Similar sequences were sought in the NCBI non-redundant sequence database (nr) and EST database (dbEST) (15) by the Blastn program and protein sequence database by the Blastx program (14) between 1 January and 31 March 1997. The 3′ sequences that did not show significant similarity to any of these entries (score <800) were selected and these clones were also sequenced from the 5′ end. These sequences were also compared with the database for any matches.

The putative ID for individual ESTs was determined by the following rules. When the Blastn score was >800, two sequences were identified as matched sequences and the hit was selected as putative identification. If the score was <300, the putative ID of the sequence was called ‘unknown’. If the score was >300 and <800 and the hit contained any of the words ‘ribosomal RNA’, ‘repeat’, ‘repetitive’, ‘vector’, ‘rRNA’ or ‘mitochondrial’ the hit was used as a putative ID. Otherwise, the putative ID of the sequence was classified as ‘unknown’. The rules were set rather empirically. For example, a score of 818 is equivalent to $P = 2.0e - 60$ for 78.00% of matches in 261 bp of sequence and a score of 819 is equivalent to $P = 6.9e - 123$ for 97.00% of matches in 171 bp of sequence. A score of 753 is equivalent to 2.40e – 118 for 98% identity in 156 bp of sequence and a score of 620 is equivalent to 5.60e – 46 for 100% identity in 124 bp of sequence.

To keep up with the rapidly growing number of ESTs, the sequence similarity searches were performed again more recently (1–22 April 1998). Thirty six percent (710/1989) of the 3′-ESTs were still ‘unknown’ in the annotated sequence database (nr) and 16% (326/1989) of them were ‘unknown’ in the EST database (dbEST).

All the information, including the cDNA and PCR primer sequences, is available through our WWW server

(http://mouse2.biosci.wayne.edu/). The EST sequence information was deposited in GenBank (2 May 1997) and has been publicly available since then through the NCBI (accession nos 1051017 and 1054202).

Gene mapping

The 751 PCR primer pairs that were designed and synthesized are available to the community through Research Genetics. Localization of genes on the mouse genetic map was performed as described with slight modifications (21). For genotyping, genomic DNAs (25 ng each) were amplified in 10 µl reactions using a Thermal Cycler Model 9600 (Perkin Elmer-Cetus). The reaction contained 10 mM Tris–HCl, pH 8.3, 50 mM KCl, 200 µM each dNTP, 1.5 mM $MgCl_2$, 0.2 µM primer (up), 0.2 µM primer (down) and 0.32 U Taq polymerase (Perkin Elmer-Cetus). The samples were heated at 95°C for 1 min to ensure complete template denaturation. Then they were amplified by 35 or 40 cycles of denaturation (at 94°C for 30 s), annealing (at 58°C for 30 s) and extension (at 72°C for 1 min), followed by 3 min extension after the last cycle. The DNAs were run on 10% non-denaturing polyacrylamide gels (20 × 15 × 0.1 cm) at 250 V for 1 h. Gels were stained with 0.5 mg/ml ethidium bromide for 30 min. Photographs were taken with a UV transilluminator (254 nm). Genotypes were scored by reading the heteroduplex bands.

The raw mapping data for the cross are accessible through the Jackson Laboratory Backcross DNA Panel Mapping Resource WWW server (http://www.jax.org/resources/documents/cmdata/BSS.html) and the Mouse Genome Database (accession no. MGD-JNUM-J41591).

Statistical analysis

Differences in the probability of appearance on each chromosome between the 7.5 EPC genes and a reference population (all coding loci) were analyzed using the χ^2 test or Fisher's exact test when appropriate. Number of all coding loci was obtained by subtracting Wsu loci and special project loci focusing on chromosomes 2 and 16 from all coding loci in the Jackson BSS Map as of 21 July 1997. Bonferroni correction was employed for exploratory analyses to avoid the multiple testing problem. We determined the new α level as 0.05/no. of comparisons (0.05/22 = 0.0023). For reference, 2504 mouse loci were collected from the Jackson Laboratory BSS Interspecific Backcross database. The StatView statistical package (Abacus Concepts, Palo Alto, CA) was used to carry out these analyses.

Gene expression analyses by RT–PCR

The same primer pairs used for gene mapping were used to amplify gene-specific PCR products from each cDNA mixture. RNA samples without reverse transcriptase treatment did not show any PCR products, indicating that there is no genomic DNA contamination in the cDNA mixtures (data not shown). Therefore, all the PCR products are derived from cDNAs and represent the expression level of the genes in the tissues. To separate the embryo proper and extraembryonic tissues including trophoblast giant cells, extraembryonic ectoderm and visceral and parietal endoderm, the 7.5 d.p.c. mouse embryos were carefully dissected out from the decidua, an incision was made in Reichert's membrane with visceral and parietal endoderm on the mesometrial side and the embryo proper was cut out. Total RNAs were extracted from each tissue sample and treated with RNase-free Dnase and reverse transcribed into cDNA.

ACKNOWLEDGEMENTS

We thank David Schlessinger for discussions and critical reading of the manuscript, Susumu Ohno, Joe Nadeau and Karen Artzt for discussion and Michael Boehnke, Morton Brown and Alan Zonderman for invaluable suggestions for statistical analysis of gene clusters. We also thank Carrie James for mouse husbandry, Sudheer Tamula and Sunil Kosuru for computational analyses. H.N. was supported by a post-doctoral fellowship from Uehara Memorial Foundation. H.F. was on leave from Daiichi Pharmaceutical Co. Ltd. The Jackson Laboratory Backcross DNA Panel Mapping Resource is supported by grant HG00941 from NHGRI. This work was mainly supported by grant HD32243 from NICHD to M.S.H.K.

ABBREVIATIONS

d.p.c., days post-conception; EPC, ectoplacental cone; ESTs, expressed sequence tags.

REFERENCES

1. Copp,A.J. (1995) Death before birth: clues from gene knockouts and mutations. *Trends Genet.*, **11**, 87–93.
2. Edmonds,D.K., Lindsay,K.S., Miller,J.F., Williamson,E. and Wood,P.J. (1982) Early embryonic mortality in women. *Fertil. Steril.*, **38**, 447–453.
3. Wilcox,A.J., Weinberg,C.R., O'Connor,J.F., Baird,D.D., Schlatterer,J.P., Canfield,R.E., Armstrong,E.G. and Nisula,B.C. (1988) Incidence of early loss of pregnancy. *N. Engl. J. Med.*, **319**, 189–194.
4. Cross,J.C., Werb,Z. and Fisher,S.J. (1994) Implantation and placenta: key pieces of the development puzzle. *Science*, **266**, 1508–1518.
5. Rossant,J. (1986) Development of extraembryonic cell lineages in the mouse embryo. In Rossant,J. and Pedersen,R.A. (eds), *Experimental Approaches to Mammalian Embryonic Development*. Cambridge University Press, Cambridge, UK, pp. 97–120.
6. Hoffman,L.H. and Wooding,F.B.P. (1993) Giant and binucleate trophoblast cells of mammals. *J. Exp. Zool.*, **266**, 559–577.
7. Barlow,D.P. (1995) Gametic imprinting in mammals. *Science*, **270**, 1610–1613.
8. Leighton,P.A., Saam,J.R., Ingram,R.S. and Tilghman,S.M. (1996) Genomic imprinting in mice: its function and mechanism. *Biol. Reprod.*, **54**, 273–278.
9. McGrath,J. and Solter,D. (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell*, **37**, 179–183.
10. Surani,M.A., Barton,S.C. and Norris,M.L. (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature*, **308**, 548–550.
11. Kajii,T. and Ohama,K. (1977) Androgenetic origin of hydatidiform mole. *Nature*, **268**, 633–634.
12. Takagi,N. and Sasaki,M. (1975) Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature*, **256**, 640–641.
13. Lyon,M.F. (1993) Epigenetic inheritance in mammals. *Trends Genet.*, **9**, 123–128.
14. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
15. Boguski,M.S., Lowe,T.M.J. and Tolstoshev,C.M. (1993) dbEST-database for 'expressed sequence tags'. *Nature Genet.*, **4**, 332–333.
16. Marra,M.A., Hillier,L. and Waterston,R.H. (1998) Expressed sequence tags—ESTablishing bridges between genomes. *Trends Genet.*, **14**, 4–7.
17. Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y. and Matsubara,K. (1992) Large scale cDNA sequencing for gene expression analysis: quantitative and qualitative aspects of gene expression in a liver cell line, Hep G2. *Nature Genet.*, **2**, 173–179.
18. Yotsumoto,S., Shimada,T., Cui,C.Y., Nakashima,H., Fujiwara,H. and Ko,M.S.H. (1998) Expression of adrenomedullin, a hypotensive peptide, in the trophoblast giant cells at the embryo implantation site in mouse. *Dev. Biol.*, in press.

19. Zhou,Y., Fisher,S.J., Janatpour,M., Genbacev,O., Dejana,E., Wheelock,M. and Damsky,C.H. (1997) Human cytotrophoblasts adopt a vascular phenotype as they differentiate. A strategy for successful endovascular invasion? *J. Clin. Invest.*, **99**, 2139–2151.
20. Risau,W. (1995) Differentiation of endothelium. *FASEB J.*, **9**, 926–933.
21. Ko,M.S.H., Wang,X., Horton,J.H., Hagen,M.D., Takahashi,N., Maezaki,Y. and Nadeau,J.H. (1994) Genetic mapping of 40 cDNA clones on the mouse genome by PCR. *Mamm. Genome*, **5**, 349–355.
22. Takahashi,N. and Ko,M.S.H. (1993) The short 3′-end region of complementary DNAs as PCR-based polymorphic markers for an expression map of the mouse genome. *Genomics*, **16**, 161–168.
23. Nagamine,C.M., Chan,K. and Lau,Y.F. (1989) A PCR artifact: generation of heteroduplexes. *Am. J. Hum. Genet.*, **45**, 337–339.
24. Rowe,L.B., Nadeau,J.H., Turner,R., Frankel,W.N., Letts,V.A., Eppig,J.T., Ko,M.S.H., Thurston,S.J. and Birkenmeier,E.H. (1994) Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. *Mamm. Genome*, **5**, 253–274.
25. Manly,K.F. (1993) A Macintosh program for storage and analysis of experimental genetic mapping data. *Mamm. Genome*, **4**, 303–313.
26. Saccone,S., Caccio,S., Perani,P., Andreozzi,L., Rapisarda,A., Motta,S. and Bernardi,G. (1997) Compositional mapping of mouse chromosomes and identification of the gene-rich regions. *Chromosome Res.*, **5**, 293–300.
27. Schuler,G.D., Boguski,M.S., Stewart,E.A., Stein,L.D., Gyapay,G., Rice,K., White,R.E., Rodriguez-Tome,P., Aggarwal,A., Bajorek,E. *et al.* (1996) A gene map of the human genome. *Science*, **274**, 540–546.
28. Bennett,D. (1975) The T-locus of the mouse. *Cell*, **6**, 441–454.
29. Silver,L.M. (1993) The peculiar journey of a selfish chromosome: mouse t haplotypes and meiotic drive. *Trends Genet.*, **9**, 250–254.
30. Ohno,S. (1973) Ancient linkage groups and frozen accidents. *Nature*, **244**, 259–262.
31. Ruddle,F.H., Bartels,J.L., Bentley,K.L., Kappen,C., Murtha,M.T. and Pendleton,J.W. (1994) Evolution of Hox genes. *Annu. Rev. Genet.*, **28**, 423–442.
32. Hardison,R. and Miller,W. (1993) Use of long sequence alignments to study the evolution and regulation of mammalian globin gene clusters. *Mol. Biol. Evol.*, **10**, 73–102.
33. Tiranti,V., Savoia,A., Forti,F., D'Apolito,M.F., Centra,M., Rocchi,M. and Zeviani,M. (1997) Identification of the gene encoding the human mitochondrial RNA polymerase (h-mtRPOL) by cyberscreening of the Expressed Sequence Tags database. *Hum. Mol. Genet.*, **6**, 615–625.
34. Hemminki,A., Markie,D., Tomlinson,I., Avizienyte,E., Roth,S., Loukola,A., Bignell,G., Warren,W., Aminoff,M., Hoglund,P., Jarvinen,H., Kristo,P., Pelin,K., Ridanpaa,M., Salovaara,R., Toro,T., Bodmer,W., Olschwang,S., Olsen,A.S., Stratton,M.R., de la Chapelle,A. and Aaltonen,L.A. (1998) A serine/threonine kinase gene defective in Peutz–Jeghers syndrome. *Nature*, **391**, 184–187.
35. Monod,J. and Jacob,F. (1961) Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.*, **3**, 318–356.
36. Vogel,F. (1969) Does the human X chromosome show evidence for clustering of genes with related functions? *J. Genet. Hum.*, **17**, 475–477.
37. Farr,C.J. and Goodfellow,P.N. (1992) Hidden messages in genetic maps. *Science*, **258**, 49.
38. Trowsdale,J. (1993) Genomic structure and function in the MHC. *Trends Genet*, **9**, 117–122.
39. Cattanach,B.M. and Jones,J. (1994) Genetic imprinting in the mouse: implications for gene regulation. *J. Inher. Metab. Dis.*, **17**, 403–420.
40. Pallavicini,A., Zimbello,R., Tiso,N., Muraro,T., Rampoldi,L., Bortoluzzi,S., Valle,G., Lanfranchi,G. and Danieli,G.A. (1997) The preliminary transcript map of a human skeletal muscle. *Hum. Mol. Genet.*, **6**, 1445–1450.
41. Spiegelman,M., Artzt,K. and Bennett,D. (1976) Embryological study of a T/t locus mutation (t^{w73}) affecting trophectoderm development. *J. Embryol. Exp. Morph.*, **36**, 373–381.
42. Silver,L.M. (1995) *Mouse Genetics: Concepts and Applications*. Oxford University Press, New York, NY, pp. 115–116.
43. Fisher,E. and Scambler,P. (1994) Human haploinsufficiency—one for sorrow, two for joy. *Nature Genet.*, **7**, 5–7.
44. McLaughlin,K.J., Szabo,P., Haegel,H. and Mann,J.R. (1996) Mouse embryos with paternal duplication of an imprinted chromosome 7 region die at midgestation and lack placental spongiotrophoblast. *Development*, **122**, 265–270.
45. Guillemot,F., Caspary,T., Tilghman,S.M., Copeland,N.G., Gilbert,D.J., Jenkins,N.A., Anderson,D.J., Joyner,A.L., Rossant,J. and Nagy,A. (1995) Genomic imprinting of Mash2, a mouse gene required for trophoblast development. *Nature Genet.*, **9**, 235–242.
46. Tada,T., Takagi,N. and Adler,I.D. (1993) Parental imprinting on the mouse X chromosome: effects on the early development of X0, XXY and XXX embryos. *Genet. Res.*, **62**, 139–148.
47. Marahrens,Y., Panning,B., Dausman,J., Strauss,W. and Jaenisch,R. (1997) Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev.*, **11**, 156–166.
48. Zechner,U., Reule,M., Orth,A., Bonhomme,F., Strack,B., Guenet, Hameister,H. and Fundele,R. (1996) An X-chromosome linked locus contributes to abnormal placental development in mouse interspecific hybrid. *Nature Genet.*, **12**, 398–403.
49. Junien,C. (1992) Beckwith–Wiedemann syndrome, tumourigenesis and imprinting. *Curr. Opin. Genet. Dev.*, **2**, 431–438.
50. Pilia,G., Hughes-Benzie,R.M., MacKenzie,A., Baybayan,P., Chen,E.Y., Huber,R., Neri,G., Cao,A., Forabosco,A. and Schlessinger,D. (1996) Mutations in GPC3, a glypican gene, cause the Simpson–Golabi–Behmel overgrowth syndrome. *Nature Genet.*, **12**, 241–247.
51. Moore,T. and Haig,D. (1991) Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet.*, **7**, 45–49.
52. Hogan,B., Beddington,R., Costantini,F. and Lacy,E. (1994) *Manipulating the Mouse Embryo: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
53. Takahashi,N. and Ko,M.S.H. (1994) Toward a whole cDNA catalog: construction of an equalized cDNA library from mouse embryos. *Genomics*, **23**, 202–210.